Differentiation of Malignant and Benign Adrenal Lesions With Delayed CT: Multivariate Analysis and Predictive Models

Chaan S. Ng[1]
Wei Wei[2]
Emre Altinmakas[3]
Xiao Li[2]
Payel Ghosh[1]
Nancy A. Perrier[4]
Elizabeth Grubbs[4]
Victor G. Prieto[5]
Jeffrey E. Lee[4]
Brian P. Hobbs[2]

Keywords: adrenal, adrenal tumors, characterization, delayed washout CT

doi.org/10.2214/AJR.17.18428

Received April 27, 2017; accepted after revision July 21, 2017.

C. S. Ng has received research funding from GE Healthcare.

Supported in part by a Cancer Center Support grant, grant P30CA016672 from the National Institutes of Health and National Cancer Institute, and the John S. Dunn, Sr. Distinguished Chair in Diagnostic Imaging.

[1]Department of Radiology, Unit 1473, University of Texas M.D. Anderson Cancer Center, 1515 Holcombe Blvd, Houston, TX 77030. Address correspondence to C. S. Ng (cng@mdanderson.org).

[2]Department of Biostatistics, University of Texas M.D. Anderson Cancer Center, Houston, TX.

[3]Department of Diagnostic Radiology, Koç University School of Medicine, Istanbul, Turkey.

[4]Department of Surgical Oncology, University of Texas M.D. Anderson Cancer Center, Houston, TX.

[5]Department of Anatomical Pathology Administration, University of Texas M.D. Anderson Cancer Center, Houston, TX.

WEB
This is a web exclusive article.

Supplemental Data
Available online at www.ajronline.org.

AJR 2018; 210:W156–W163

0361–803X/18/2104–W156

OBJECTIVE. The purpose of this study is to identify imaging and patient parameters that affect the diagnostic performance of delayed contrast-enhanced CT for distinguishing malignant from benign adrenal lesions larger than 1 cm in adult patients and to derive predictive models.

MATERIALS AND METHODS. This retrospective study assessed 97 pathologically proven adrenal lesions that had undergone unenhanced, portal venous, and 15-minute delayed CT. Quantitatively, single-parameter evaluations of lesion attenuation (in Hounsfield units) and absolute percentage enhancement washout (APEW) and relative percentage enhancement washout (RPEW) were performed. In addition, descriptive CT features (lesion size, margin definition, heterogeneity vs homogeneity, fat, and calcification) and patients' demographic characteristics and medical history of malignancy were evaluated for association with lesion status using multiple logistic regression with stepwise model selection. Areas under the ROC curve (A_z) were determined for univariate and multivariate analyses. Leave-one-lesion-out cross-validation was applied to ascertain the predictive performance of single-parameter and multivariate evaluations.

RESULTS. The A_z values for unenhanced attenuation, portal venous attenuation, delayed attenuation, APEW, and RPEW were 0.835, 0.534, 0.847, 0.792, and 0.871, respectively. Multivariate analyses revealed that portal venous attenuation, delayed attenuation, and APEW were significant features, with an A_z of 0.923 when combined. The addition of the descriptive CT features increased the A_z to 0.938; patient age and a history of malignancy were additional significant factors, increasing the A_z to 0.956 and 0.972, respectively. The combined predictive classifier yielded 89% accuracy under cross-validation, compared with the best commonly applied single-parameter evaluation (77% for RPEW < 40%).

CONCLUSION. Multivariate imaging evaluation applied to delayed contrast-enhanced CT alone, with or without patient characteristics, improves diagnostic performance for characterizing adrenal lesions beyond those of single-parameter evaluations. Predictive formulas assessing the probabilities of lesion benignity or malignancy are provided.

The detection of an adrenal lesion presents a substantial challenge for radiologists, clinicians, and their patients. The challenge centers on the need to differentiate between malignant and benign lesions; the former have grave implications, whereas the latter are much more common [1]. Adrenal lesions may be encountered in a variety of clinical settings, such as in the course of cancer staging, or may be entirely incidental. Incidentally detected adrenal lesions have been reported in up to 9% of abdominal CT scans [2, 3].

Whatever the clinical scenario in which a lesion is detected, a major component of many clinical algorithms designed to assist in making the fundamental differentiation between malignancy and benignity is the use of adrenal protocol CT. Adrenal protocol CT scans use delayed contrast-enhanced imaging in which lesion washout after IV contrast medium administration is evaluated. Quantitatively, washout characteristics are based on absolute percentage enhancement washout (APEW) or relative percentage enhancement washout (RPEW), depending on the availability or not of associated unenhanced images, respectively [4–10].

The utility of washout characteristics on CT is based on the observation that adenomas (both lipid poor and lipid rich) show more-rapid washout of IV contrast agent than

do nonadenomas [4, 5, 7]. Different washout delay intervals have been investigated, but 10–15 minutes is widely regarded as a reliable and practical delay interval [5, 8, 10, 11].

A variety of descriptive radiologic features, such as lesion margin and internal heterogeneity, have also been used to attempt to differentiate malignancy from benignity [12, 13]. In routine clinical radiologic practice, additional information, such as patient demographics and medical history [14], may or may not also be available.

On the basis of previous CT studies using quantitative evaluations of adrenal protocol CT images, a variety of cutoff thresholds for the various aforementioned individual CT characteristics have been suggested to differentiate benign from malignant lesions, with associated sensitivities and specificities. Clinical algorithms and pathways that incorporate some of these concepts and thresholds have been suggested to assist in the practical management of adrenal lesions [15, 16]. However, an analysis of the interaction and appropriate combination of these single-parameter evaluations (APEW, RPEW, and individual CT attenuations of each phase) has not been undertaken for imaging data obtained from adrenal protocol CT scans alone, on which substantial management decisions frequently rest.

The commonly used cutoff method mentioned already does not convey an understanding of the likelihood or probability of malignancy or benignity; instead, lesion classification is based on whether the parameter of interest is simply above or below the cutoff value. Furthermore, to our knowledge, there has not been an evaluation of the extent to which tissue characterization may be enhanced by integrating multiple quantitative features with other clinical factors that might contribute to overall diagnostic performance. Through integration of factors that confer malignancy risk, whether quantitative, descriptive, demographic, or history, such an approach allows the development of multivariate models that facilitate multifeature evaluations for diagnostic decision making based on probability.

The objective of our study was to investigate, for adult patients with an adrenal lesion larger than 1 cm who underwent adrenal protocol CT with delayed washout, which imaging features influence the diagnostic performance for distinguishing malignant from benign adrenal lesions in multivariate analysis and to evaluate their joint predictive capabilities on the basis of imaging alone, as well as when combined with available patient demographic and clinical characteristics. Predictive models assessing the probability of benignity or malignancy using quantitative data from adrenal protocol CT images alone, with or without additional clinical data, are also presented.

Materials and Methods

This retrospective study was approved by the University of Texas M.D. Anderson Cancer Center institutional review board. A waiver of informed consent was granted.

Patients

A search of the institutional pathology database was undertaken for the term "adrenal," looking for pathologic findings and rendered diagnoses between January 2001 and January 2010. Patients who had CT images available in our radiology PACS before their pathologic diagnosis date were then identified.

Inclusion criteria for our study were age older than 18 years at the time of CT; no clinical or biochemical evidence of functioning adrenal lesions (because these typically undergo alternative focused clinical evaluations); interval between pathologic analysis and precedent CT scan less than 160 days; CT images consisting of unenhanced, portal venous phase, and delayed phase contrast-enhanced images (required to be > 9 minutes after the portal venous phase); and size of adrenal lesion greater than 1 cm and larger than twice the CT image slice thickness. The pathology database yielded 792 cases; in 332 cases, there was only normal adrenocortical tissue with no pathologic abnormality. Of the remaining 460 cases, 97 lesions (96 patients) met our inclusion criteria.

CT Scan

A range of CT scanners had been used, with the following parameters: median tube current, 265 mA (range, 86–630 mA); tube voltage, 120 kVp in 96 cases; and median slice thickness, 2.5 mm (range, 2.5–5.0 mm). Scanner models were GE Healthcare LightSpeed 16 (n = 51), GE Healthcare LightSpeed VCT (n = 27), GE Healthcare LightSpeed Plus (n = 8), GE Healthcare LightSpeed QX/I (n = 8), and other (n = 3).

Contrast-enhanced CT scans had been obtained 60–70 seconds after IV administration of 100–150 mL of nonionic contrast agent (iohexol, 300 mg I/mL; GE Healthcare) at a rate of 2.0–3.0 mL/s by use of a power injector. The median time delay between the portal venous phase and delayed phase acquisitions was 14.1 minutes (range, 9.0–19.3 minutes).

CT Analysis

Both quantitative and descriptive evaluations of the adrenal lesions were noted on our PACS (IntelliSpace PACS Radiology, version 4.4, Philips Healthcare). Images were reviewed on soft-tissue windows (width, 400 HU; level, 50 HU), by a radiologist with more than 5 years' postresidency experience in abdominal CT who was blinded to the lesion's status.

Quantitative CT evaluation—For each study, we selected as index CT image the CT slice that contained the maximal axial cross-sectional area of the adrenal mass. An ROI was carefully drawn freehand, with an electronic cursor and mouse, around the periphery of the adrenal lesion on the CT image that contained the maximal cross-sectional area. The extreme edges of the mass were meticulously avoided to prevent potential partial volume artifacts. The ROI was saved and transposed onto the image of the two other CT series containing the maximal cross-sectional area of the target lesion; translational adjustments to the ROI were made as necessary to correct any axial misalignments in any given image, but the shape and size of the original ROI was preserved. This was undertaken within a graphic user interface in Matlab (version 2013b, MathWorks) interfacing the National Institutes of Health ImageJ software [17]. The CT attenuation (in Hounsfield units) for each pixel within each ROI, for each CT phase, were saved and exported for statistical analysis.

APEW and RPEW characteristics of lesions were calculated on the basis of the pixel values of respective ROIs, as follows: APEW = [(portal venous – delayed) / (portal venous – unenhanced)] × 100, and RPEW = [(portal venous – delayed) / portal venous] × 100, where unenhanced, portal venous, and delayed are the mean attenuations on unenhanced, portal venous, and delayed phase contrast-enhanced CT, respectively [7].

Descriptive CT evaluation—The following descriptive CT features of the whole adrenal lesion were assessed: lesion size (maximal long- and short-axis diameters); lesion margin, whether ill- or well-defined; lesion texture, whether homogeneous or heterogeneous on contrast-enhanced images; calcification, whether present or absent; and fat, whether present (less than –20 HU) or absent.

Demographics and Medical History

Patient age and sex were obtained from the patient medical records. The presence or absence of a medical history of malignancy before the date of CT was recorded from thorough review of patient electronic medical records.

The study cohort consisted of 97 adrenal lesions in 96 patients (45 men [median age, 55 years; range, 27–82 years] and 51 women [median age, 51 years;

range, 24–81 years]). One patient had bilateral lesions. Of the 97 lesions, 75 (77.3%) were benign and 22 (22.7%) were malignant. The malignant diagnoses were metastasis (n = 15), adrenocortical carcinoma (n = 3), lymphoma (n = 2), and unclassified malignant neoplasm (n = 2). The benign diagnoses were adenoma (n = 68), hyperplasia (n = 3), myelolipoma (n = 2), hemorrhage (n = 1), and unclassified benign neoplasm (n = 1). Regarding the benign lesions, of the 68 adenomas, 21 (30.9%) were greater than or equal to 4 cm, and 27 (39.7%) were greater than or equal to 3.5 cm. The two myelolipomas in the cohort were larger than 8 cm; one was resected for clinical and symptomatic reasons, and the other was biopsied because of clinical concern for a liposarcoma. The hemorrhagic lesion was 6 cm and underwent surgery because of patient concern. Pathologic diagnoses were established by adrenalectomy (n = 88), biopsy (n = 7), and fine-needle aspiration (n = 2). Fifty-one lesions were on the left side and 46 were on the right side. The median interval between the CT scan and surgery or biopsy was 20 days (range, 1–124 days).

At the time of the index CT, the existence of a prior primary malignancy was known in 27 of the 97 lesions (27.8%), and in the remaining 70, there was no known previous malignancy.

TABLE 1: Univariate Analyses

Available Data, Parameter	Comparison	Odds Ratio (Malignancy/Benignity)	p[a]
CT quantitative factors			
Unenhanced phase attenuation (HU)	1 Unit increase	1.10	**< 0.0001**
Portal venous phase attenuation (HU)	1 Unit increase	1.00	0.72
Delayed phase attenuation (HU)	1 Unit increase	1.08	**< 0.0001**
RPEW (ratio)	1 Unit increase	0.95	**< 0.0001**
APEW (ratio)	1 Unit increase	0.99	0.15
CT descriptive factors			
Long-axis diameter (cm)	1 Unit increase	1.20	**0.01**
Short-axis diameter (cm)	1 Unit increase	1.18	0.06
Calcification	Yes vs no	0.97	0.97
Fat	Yes vs no	0.40	0.40
Texture	Homogeneous vs heterogeneous	0.39	0.07
Demographic factors			
Age (y)	1 Unit increase	1.05	**0.02**
Sex	Male vs female	0.95	0.92
Medical history of malignancy	Yes vs no	8.35	**0.0001**

Note—RPEW = relative percentage enhancement washout, APEW = absolute percentage enhancement washout.
[a]Bold type indicates statistical significance.

Statistical Analysis

Logistic regression models were used to identify patient-level imaging and clinical features for which associations with pathologic diagnosis (malignant vs benign) were evident, as well as to evaluate their significance in multivariate analysis. Initially, univariate logistic regression was used to identify associations between pathologic diagnosis and individual features obtained from evaluating the following four types of patient-level information: quantitative CT evaluation, which was based on measurements of lesion CT attenuation; descriptive CT evaluation, which was based on assessment of lesion characteristics, as already described; patient demographics, including patient age and sex; and medical history (i.e., of malignancy). Holm-Bonferroni correction was applied to control overall type I error rate at 5% among all univariate hypothesis tests conducted within each of the four feature classes [18].

Thereafter, several multiple logistic regression models were fit sequentially to reflect increasing levels of information that might be available in routine clinical radiologic practice. Specifically, multiple quantitative CT features were evaluated first for association with lesion status and then were combined successively with descriptive features, demographics, and medical history. Final models were selected using backward elimination based on the likelihood ratio test. An ROC curve was fitted to the linear predictors that resulted from estimation of each logistic regression model. For each model, the area under the ROC curve (A_z) was estimated along with DeLong 95% CI [19]. Partial effects were evaluated for significance using likelihood ratio tests.

After identifying features that exhibited evidence of association with pathologic diagnosis, the predictive accuracy of multifeature evaluations was ascertained. A final predictive classifier was determined for each level of source information using logistic regression with weighted estimation. Gaussian density approximation was applied to the sample of multivariate continuous features and was used to assign a weight to each observation reflecting its relative population frequency. The resultant weighted regression coefficients are reported for each type of predictive classifier. By assigning each lesion to the tissue class, resulting in higher predicted probability, diagnostic accuracy is reported on the basis of the full data as well as under a leave-one-lesion-out cross-validation strategy that adjusts the measure for predictive loss due to sampling variability [20, 21].

With 97 lesions and 22 malignancies, the study provides at least 80% power to detect a difference of 0.17 between the A_z under the null hypothesis of 0.5 and an A_z under the alternative hypothesis of 0.67 using a one-sided z-test at significance level of 0.05. All statistical analyses were performed using SAS (version 9, SAS Institute) or R (version 3.1.2, R Development Core Team) statistical software.

Results

CT Lesion Characteristics and Univariate Analyses

Summaries of the quantitative and descriptive CT lesion characteristics and medical history data are presented in Tables S1 and S2, which can be viewed in the *AJR* electronic supplement to this article (available at www.ajronline.org). For our quantitative CT parameters, unenhanced attenuation was significantly higher, and RPEW and APEW were significantly lower, for malignant than benign lesions ($p < 0.0001$).

On univariate analyses, the quantitative CT parameters unenhanced and delayed attenuation were significantly associated with malignancy ($p < 0.0001$; Table 1), with an increasing risk of malignancy for larger values of the unenhanced and delayed phases. The derived parameter RPEW (but not APEW) was also significantly associated with malignancy ($p < 0.0001$), with larger RPEW decreasing the risk of malignancy. In terms of diagnostic performance, unenhanced, portal venous, and delayed phases yielded A_z values of 0.835, 0.534, and 0.847, respectively; the

TABLE 2: Diagnostic Performance of Individual CT Quantitative Parameters and Predictive Accuracy of Commonly Used Single-Parameter Thresholds

Parameter	A_z (95% CL)	Accuracy of Single-Parameter Thresholds		
		Cutoff Threshold	Accuracy (%)	Predictive Accuracy (Model Prediction) (%)
Unenhanced phase attenuation	0.835 (0.758–0.911)	> 10 HU	54.6	74.2
Portal venous phase attenuation	0.534 (0.500–0.665)			
Delayed phase attenuation	0.847 (0.762–0.932)			
RPEW	0.871 (0.802–0.940)	< 40%	78.4	77.3
APEW	0.792 (0.706–0.878)	< 60%	69.1	72.2

Note—A_z = area under the ROC curve, CL = confidence limit, RPEW = relative percentage enhancement washout, APEW = absolute percentage enhancement washout.

A_z for RPEW (0.871) was higher than that for APEW (0.792) (Table 2 and Fig. 1).

Lesion size was borderline associated with risk of malignancy ($p \leq 0.06$), but none of the other descriptive CT features were associated with malignancy (Table 1). Regarding demographic features, age was significantly associated with malignancy ($p = 0.02$). A history of malignancy was highly associated with an increased risk of malignancy of the adrenal lesion ($p = 0.0001$; Table 1).

Multivariate Analyses

Our multivariate inference of the quantitative CT features indicated that APEW, portal venous attenuation, and delayed attenuation each contributed significant partial effects in multiple regression analysis. Incorporating these three features in the logistic model yielded an A_z of 0.923 (Table 3 and Fig. 2), which was higher than that obtained for any individual feature in univariate analysis.

With the addition of descriptive CT features, long-axis lesion diameter was found to contribute significantly as an additional factor, when integrated with the significant quantitative features; combining these four CT features yielded an A_z of 0.938. No other descriptive features contributed significant partial effects.

With the sequential addition of patient demographics and medical history, age was an additional significant factor, yielding an A_z of 0.956, with knowledge of a prior malignancy yet a further significant factor, increasing A_z to 0.972.

Predictive Power

The analyses already discussed were implemented using a conventional hypothesis testing framework with the intention of describing the associations that were evident in our study. We also evaluated the predictive power of integrating the CT parameters without or with the clinical and demographic considerations to ascertain the extent to which multivariate evaluations can be used to accurately classify, or correctly predict, future lesions of unknown status on the basis of our study. The resultant classifiers attained predictive accuracies of 85–89% under leave-one-lesion-out cross-validation, and 87–93% when applied to the entire sample (Table 3). These are higher than the predictive accuracies attained when applying commonly used cutoffs for the single-parameter evaluations (unenhanced attenuation > 10 HU, APEW < 40%, and RPEW < 60%), which yielded predictive accuracies of only 72–77% (Table 2).

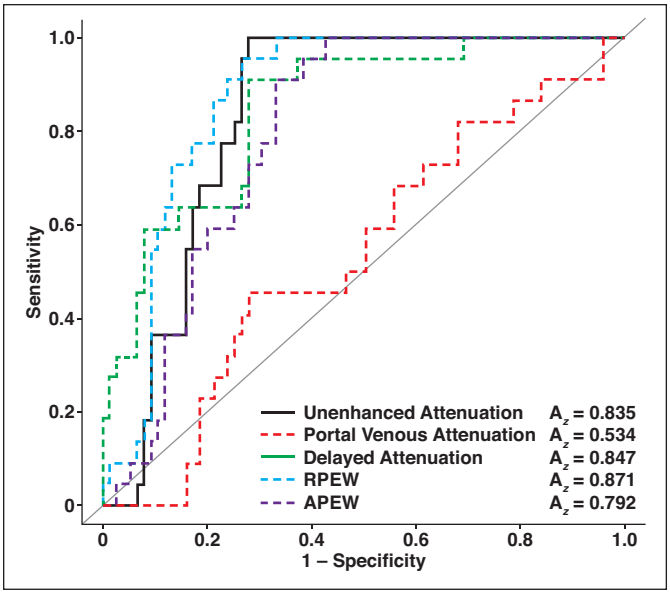

Fig. 1—Summary ROC plot for univariate analyses. A_z = area under the ROC curve, RPEW = relative percentage enhancement washout, APEW = absolute percentage enhancement washout.

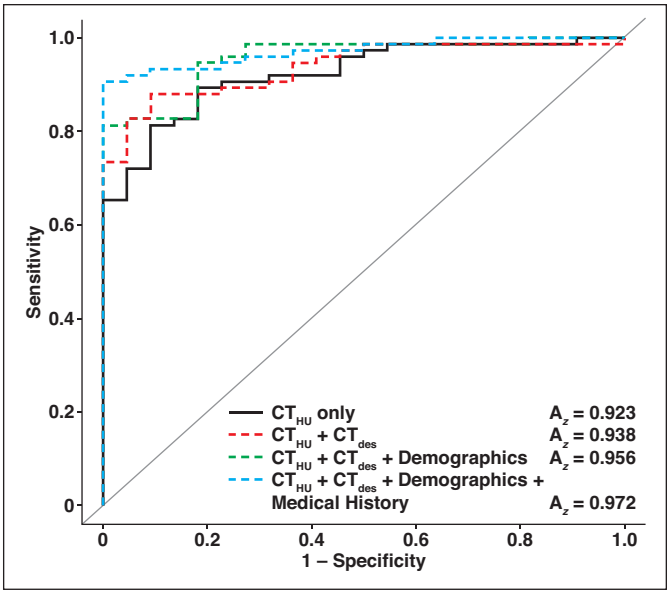

Fig. 2—Summary ROC plot for multivariate analyses. Multivariate models were based on increasing levels of available information. A_z = area under the ROC curve, CT_{HU} = quantitative CT data in Hounsfield units (HU), CT_{des} = descriptive CT features.

TABLE 3: Multivariate Analyses, Associated Diagnostic Performance of Model, and Predictive Accuracy of Final Classifiers

Model, Covariates	Comparison	Odds Ratio (Malignancy/Benignity)	*p*	A_z (95% CI)	Predictive Accuracy (%)	
					Applying Full Data	Applying Leave-One-Lesion-Out Cross-Validation
CT quantitative factors						
Portal venous phase attenuation (HU)	1 Unit increase	0.90	< 0.0001	0.923 (0.870–0.976)	89	85
Delayed phase attenuation (HU)	1 Unit increase	1.22	< 0.0001			
APEW (%)	1 Unit increase	1.04	0.012			
CT quantitative + CT descriptive						
Portal venous phase attenuation (HU)	1 Unit increase	0.91	0.0005	0.938 (0.892–0.983)	87	85
Delayed phase attenuation (HU)	1 Unit increase	1.22	< 0.0001			
APEW (%)	1 Unit increase	1.04	0.012			
Long-axis diameter (cm)	1 Unit increase	1.24	0.04			
CT quantitative + CT descriptive + demographic						
Portal venous phase attenuation (HU)	1 Unit increase	0.91	0.0006	0.956 (0.919–0.994)	92	86
Delayed phase attenuation (HU)	1 Unit increase	1.24	< 0.0001			
APEW (%)	1 Unit increase	1.06	0.005			
Long-axis diameter (cm)	1 Unit increase	1.45	0.006			
Age (y)	1 Unit increase	1.11	0.007			
CT quantitative + CT descriptive + demographic + medical history						
Portal venous phase attenuation (HU)	1 Unit increase	0.91	0.017	0.972 (0.944–0.999)	93	89
Delayed phase attenuation (HU)	1 Unit increase	1.28	< 0.0001			
APEW (%)	1 Unit increase	1.07	0.002			
Long-axis diameter (cm)	1 Unit increase	1.93	0.0002			
Age (y)	1 Unit increase	1.10	0.03			
Medical history of malignancy	Yes vs No	> 10	0.0008			

Note—A_z = area under the ROC curve, APEW = absolute percentage enhancement washout.

In Table 4, we present the regression coefficients for the predictive classifiers according to the data available. The estimated probability (π) of malignancy for a lesion can be obtained from the predictor variables (X_1 = portal venous phase attenuation, X_2 = delayed phase attenuation, X_3 = APEW, X_4 = long-axis diameter, X_5 = age, and X_6 = medical history) and the regression coefficients (a, b, c, d, e, f, and g) provided in Table 4, by Equation 1, as follows:

$$\pi(X_1,X_2,X_3,X_4,X_5,X_6) = \frac{1}{1+\exp\{-(a+X_1b+X_2c+X_3d+X_4e+X_5f+X_6g)\}}, \quad (1)$$

where exp is the exponential function. Examples of how to use the formula are presented in Appendix S3, which can be viewed on the *AJR* website (www.ajronline.org).

Discussion

The washout characteristics of adrenal lesions are widely used in clinical practice as a method of differentiating benign from malignant lesions. In our study, univariate analyses indicated that the washout parameter RPEW

TABLE 4: Regression Coefficients for Predicting the Probability of Malignancy of an Adrenal Lesion

Model	Factor (Variable)						
	Intercept (*a*)	Portal venous Phase Attenuation (*b*)	Delayed Phase Attenuation (*c*)	Absolute Percentage Enhancement Washout (*d*)	Long-Axis Diameter (*e*)	Age (*f*)	Medical History (*g*)
CT quantitative	–7.7673	–0.1458	0.2807	0.0765	0	0	0
CT quantitative + CT descriptive	–8.3468	–0.1198	0.2589	0.0627	0.0848	0	0
CT quantitative + CT descriptive + demographic	–12.5632	–0.1282	0.2637	0.0746	0.1730	0.0658	0
CT quantitative + CT descriptive + demographic + medical history	–25.7660	–0.2140	0.4539	0.1446	0.3278	0.1193	4.6300

Note—On the basis of our study data and analysis strategy, the estimated probability (π) of observing a malignant lesion can be obtained from Equation 1, as shown in the Materials and Methods section, which is based on the linear combination of the regression coefficients provided here and observed predictor variables.

was significantly associated with lesion status ($p < 0.0001$), with A_z of 0.871; APEW was not significant and had a lower A_z (0.792).

Previous studies have reported excellent performance characteristics when using delayed imaging. Korobkin et al. [4] have reported an A_z of 0.99 with 1-hour delayed images (39 patients; 51 lesions, of which two had pathology confirmation). Boland et al. [5] have reported with delayed scans at 12–18 minutes, an A_z of 0.986 (46 lesions, of which 12 had pathologic confirmation). Blake et al. [22] have reported an A_z of 0.892 for APEW and 0.985 for RPEW using delayed scanning at 10 minutes (99 patients; 112 lesions, of which 28 had pathology confirmation). These previously reported A_z performance characteristics are higher than our current results. One possible contributing factor may be that, unlike these previous reports, in our study the reference standard for classification of all lesions was based on pathologic diagnoses; the importance of this has been suggested previously [23]. Furthermore, it should also be noted that classifications and case mix of lesions differ between studies; for example, some studies exclude myelolipomas, and others assess for nonadenomas, which may include pheochromocytomas and myelolipomas.

Our univariate analyses suggested that there were other quantitative radiologic and descriptive radiologic features that were significantly associated with malignancy, particularly increases in unenhanced attenuation ($p < 0.0001$) and lesion size ($p = 0.01$). Unenhanced attenuation is a widely used parameter in clinical practice; in our study, it had an A_z of 0.835. Some studies have reported better performance for this feature than ours; for example, Blake et al. [22] have reported A_z of 0.912. Interestingly, these authors reported the same relative performance of the quantitative CT parameters as we observed, namely, in order of increasing A_z (APEW, unenhanced attenuation, and RPEW).

Additional patient characteristics conferring malignancy risk are often, but not always, available at the time of diagnostic evaluation. In our study, the presence of a history of malignancy was significantly associated with malignancy of the adrenal lesion ($p = 0.0001$). Some previous studies have examined the associations of a variety of factors on lesion classification [14, 24–26], but they have not analyzed their independent contributions to diagnostic performance. We undertook multivariate analyses to identify the imaging features that jointly influence the diagnostic performance of CT and to characterize their relative contributions to predicting a patient's malignancy status when considered conjointly with several nonradiological characteristics available in our study. Our analyses were undertaken in an incremental fashion to reflect the increasing levels of information that may be available in day-to-day clinical practice, including quantitative radiologic lesion characteristics (CT attenuation and washout), descriptive radiologic lesion characteristics (morphologic features), patient demographics (age and sex), and availability of medical history (of malignancy).

Our multivariate analyses of quantitative radiologic characteristics showed that APEW and portal venous phase and delayed phase attenuation contributed significant partial effects. When used in combination, these parameters yielded an A_z of 0.923, which was superior to the A_z of the individual covariates. Of note, the contribution of unenhanced attenuation is embedded in the parameter APEW. Combining data derived from unenhanced attenuation and washout data has been explored in a few studies, such as in the aforementioned study by Blake et al. [22] and a study by Caoili et al. [27] (166 lesions; 46 with pathologic confirmation). These studies essentially applied discrete cutoff thresholds in a stepwise manner, without examining their independent contributions in multivariate analysis; nevertheless, they found improvements in diagnostic performance.

When descriptive radiologic characteristics were added to the multivariate model, we found that lesion size was a significant, and in fact the only significant, covariate from this category, and increased the A_z to 0.938. Multivariate analyses showed that each successive level of available information increased diagnostic performance; the addition of age increased the A_z to 0.956 and, subsequently, with the addition of history of malignancy, the A_z increased to 0.972. Of note, with increasing levels of information, the same preceding quantitative and descriptive covariates remained significant factors in the models, suggesting their robustness. Conversely, a consistent set of covariates were found not to contribute significantly in multivariate modeling.

Through application of our two-stage inferential approach, we have identified the significant covariates that contribute to characterization of lesions using descriptive analysis (based on regression) and have also ascertained the predictive power of multivariate evaluations that integrate the resultant significant features using multiple regression. Binary classifiers are often univariate and rely on cutoffs, which fail to convey the likelihood or probability of malignancy or benignity. In contrast, a binary classifier was obtained in our study through statistical modeling by assigning each target lesion to the candidate class that resulted in higher estimated probability. The approach yielded 89% accuracy under cross-validation for predicting malignant from benign lesions on the basis of the full model. In comparison, the highest predictive accuracy attained from single-parameter evaluations using commonly applied thresholds was 77% (which was obtained with the parameter RPEW < 40%).

Our study focused on data related specifically to CT scans with delayed washout images. As such, the modeling relates to patients who have reached the point in the various clinical algorithms that they then undergo adrenal protocol CT. Our work is designed to assess the probabilities of malignancy or benignity on the basis of quantitative CT data obtained from that imaging session alone, combined with other descriptive imaging, demographic information, or medical history that might be available. In current clinical practice, when faced with adrenal protocol CT images, assessment of benignity or malignancy of a lesion is based on whether its RPEW or APEW is above or below the respective threshold values. This threshold approach does not convey any sense of the likelihood or probability of the classification. Indeed, proximity or otherwise of the values to the threshold have no bearing; values are simply either above or below the cutoff. It is also conceivable that determinations based on RPEW and APEW may be discordant. In comparison, from our multivariate analyses, we are able to present predictive models that yield probabilities of malignancy or benignity of lesions according to the various levels of data that may be available (see Table 4 and the online data supplement). Furthermore, the predictive accuracies using the multivariate models (85–89%) are higher than when using the threshold approaches (72–77%).

We recognize and acknowledge several possible limitations in our study. This was a retrospective study and thus subject to potential biases, such as in the selection of patients who ultimately had a pathologic diagnosis (i.e., surgery or biopsy); however, it had

the benefit that all diagnoses were histologically confirmed. It could be considered surprising that there were lesions in our cohort with low unenhanced CT attenuation (< 10 HU), namely that they had undergone biopsy or resection; however, it should be borne in mind that there are definitely clinical and patient circumstances that can supervene imaging findings, for example, when the finding of malignancy or benignity in the lesion might significantly alter the direction of management. Overall there was a preponderance of benign lesions, which possibly introduces some bias; however, the rigor of this study required us to only include lesions with pathologic diagnoses, and this is what ultimately dictated the distribution of diagnoses.

The study was limited in size, but nevertheless significant findings were evident. The study spanned several years and used a range of CT scanners and CT protocols, which was inevitable to gain the number of histologically confirmed lesions. We recognize that different scanners and acquisition protocols, including a range of delay times, have potential effects on attenuation measurements [28–31], but assessment of these was beyond the scope of our study; it arguably improves the generalizability of the results. We did not assess or include in our modeling changes in lesion size or the specific previous primary malignancy, because we had insufficient data for satisfactory modeling. We did not assess observer variability [32]. We recognize that there may have been institutional biases that influenced the management paths of patients, such as those who underwent adrenal protocol CT and those selected for biopsy or surgery [15, 16]; many factors come into play in these decisions, including institutional workflows, the degree of certainty required in the context of oncologic staging (which itself can vary widely according to the overall findings, protocols, and clinicians) to determine therapies, and patient anxiety. We recognize that our statistical analyses and resultant model parameters have been derived from a specific cohort of patients, and it requires independent validation using other cohorts. Nevertheless, the analytical methods and concepts of predictive modeling that we have presented can still be used.

In conclusion, our multivariate analyses have identified significant imaging features among quantitative CT information that is commonly collected and used from delayed adrenal CT (lesion attenuation and washout) for differentiating malignant from benign lesions; their combination enhances the diagnostic performance beyond single-parameter evaluations. Diagnostic performance increases further with additional clinical information, specifically, lesion size, patient age, and presence or absence of previous malignancy; their combination enables the development of probabilistic models for differentiating malignant from benign adrenal lesions. On the basis of these models, we provide formulas whose covariates vary according to the available data, which yield predictive probabilities of benignity or malignancy of adrenal lesions.

References

1. Mazzaglia PJ. Radiographic evaluation of nonfunctioning adrenal neoplasms. *Surg Clin North Am* 2014; 94:625–642
2. Glazer HS, Weyman PJ, Sagel SS, Levitt RG, McClennan BL. Nonfunctioning adrenal masses: incidental discovery on computed tomography. *AJR* 1982; 139:81–85
3. Dunnick NR, Korobkin M, Francis I. Adrenal radiology: distinguishing benign from malignant adrenal masses. *AJR* 1996; 167:861–867
4. Korobkin M, Brodeur FJ, Francis IR, Quint LE, Dunnick NR, Goodsitt M. Delayed enhanced CT for differentiation of benign from malignant adrenal masses. *Radiology* 1996; 200:737–742
5. Boland GW, Hahn PF, Pena C, Mueller PR. Adrenal masses: characterization with delayed contrast-enhanced CT. *Radiology* 1997; 202:693–696
6. Szolar DH, Kammerhuber F. Quantitative CT evaluation of adrenal gland masses: a step forward in the differentiation between adenomas and nonadenomas? *Radiology* 1997; 202:517–521
7. Szolar DH, Kammerhuber FH. Adrenal adenomas and nonadenomas: assessment of washout at delayed contrast-enhanced CT. *Radiology* 1998; 207:369–375
8. Korobkin M, Brodeur FJ, Francis IR, Quint LE, Dunnick NR, Londy F. CT time-attenuation washout curves of adrenal adenomas and nonadenomas. *AJR* 1998; 170:747–752
9. Peña CS, Boland GW, Hahn PF, Lee MJ, Mueller PR. Characterization of indeterminate (lipid-poor) adrenal masses: use of washout characteristics at contrast-enhanced CT. *Radiology* 2000; 217:798–802
10. Caoili EM, Korobkin M, Francis IR, Cohan RH, Dunnick NR. Delayed enhanced CT of lipid-poor adrenal adenomas. *AJR* 2000; 175:1411–1415
11. Park SW, Kim TN, Yoon JH, et al. The washout rate on the delayed CT image as a diagnostic tool for adrenal adenoma verified by pathology: a multicenter study. *Int Urol Nephrol* 2012; 44:1397–1402
12. Gufler H, Eichner G, Grossmann A, et al. Differentiation of adrenal adenomas from metastases with unenhanced computed tomography. *J Comput Assist Tomogr* 2004; 28:818–822
13. Song JH, Grand DJ, Beland MD, Chang KJ, Machan JT, Mayo-Smith WW. Morphologic features of 211 adrenal masses at initial contrast-enhanced CT: can we differentiate benign from malignant lesions using imaging features alone? *AJR* 2013; 201:1248–1253
14. Song JH, Chaudhry FS, Mayo-Smith WW. The incidental adrenal mass on CT: prevalence of adrenal disease in 1,049 consecutive adrenal masses in patients with no known malignancy. *AJR* 2008; 190:1163–1168
15. Boland GW, Blake MA, Hahn PF, Mayo-Smith WW. Incidental adrenal lesions: principles, techniques, and algorithms for imaging characterization. *Radiology* 2008; 249:756–775
16. Berland LL, Silverman SG, Gore RM, et al. Managing incidental findings on abdominal CT: white paper of the ACR incidental findings committee. *J Am Coll Radiol* 2010; 7:754–773
17. Schneider CA, Rasband WS, Eliceiri KW. NIH Image to ImageJ: 25 years of image analysis. *Nat Methods* 2012; 9:671–675
18. Proschan MA, Waclawiw MA. Practical guidelines for multiplicity adjustment in clinical trials. *Control Clin Trials* 2000; 21:527–539
19. DeLong ER, DeLong DM, Clarke-Pearson DL. Comparing the areas under two or more correlated receiver operating characteristic curves: a nonparametric approach. *Biometrics* 1988; 44:837–845
20. Picard RR, Cook RD. Cross-validation of regression models. *J Am Stat Assoc* 1984; 79:575–583
21. Geisser S. *Predictive inference*. Oxford, UK: Taylor & Francis, 1993
22. Blake MA, Kalra MK, Sweeney AT, et al. Distinguishing benign from malignant adrenal masses: multi-detector row CT protocol with 10-minute delay. *Radiology* 2006; 238:578–585
23. Park BK, Kim B, Ko K, Jeong SY, Kwon GY. Adrenal masses falsely diagnosed as adenomas on unenhanced and delayed contrast-enhanced computed tomography: pathological correlation. *Eur Radiol* 2006; 16:642–647 [Erratum in *Eur Radiol* 2006; 16:768]
24. Park SY, Park BK, Park JJ, Kim CK. CT sensitivities for large (>=3 cm) adrenal adenoma and cortical carcinoma. *Abdom Imaging* 2015; 40:310–317
25. Azoury SC, Nagarajan N, Young A, et al. Computed tomography in the management of adrenal tumors: does size still matter? *J Comput Assist Tomogr* 2017; 41:628–632
26. Foo E, Turner R, Wang KC, et al. Predicting malignancy in adrenal incidentaloma and evaluation of a novel risk stratification algorithm. *ANZ J*

Surg 2017 Jan 24 [Epub ahead of print]

27. Caoili EM, Korobkin M, Francis IR, et al. Adrenal masses: characterization with combined unenhanced and delayed enhanced CT. *Radiology* 2002; 222:629–633
28. Bae KT, Fuangtharnthip P, Prasad SR, Joe BN, Heiken JP. Adrenal masses: CT characterization with histogram analysis method. *Radiology* 2003; 228:735–742
29. Jhaveri KS, Wong F, Ghai S, Haider MA. Comparison of CT histogram analysis and chemical shift MRI in the characterization of indeterminate adrenal nodules. *AJR* 2006; 187:1303–1308
30. Hammarstedt L, Thilander-Klang A, Muth A, Wangberg B, Oden A, Hellstrom M. Adrenal lesions: variability in attenuation over time, between scanners, and between observers. *Acta Radiol* 2013; 54:817–826
31. Lamba R, McGahan JP, Corwin MT, et al. CT Hounsfield numbers of soft tissues on unenhanced abdominal CT scans: variability between two different manufacturers' MDCT scanners. *AJR* 2014; 203:1013–1020
32. Remer EM, Motta-Ramirez GA, Shepardson LB, Hamrahian AH, Herts BR. CT histogram analysis in pathologically proven adrenal masses. *AJR* 2006; 187:191–196

FOR YOUR INFORMATION

The data supplement accompanying this web exclusive article can be viewed by clicking "Supplemental" at the top of the article.